Algebra and Calculus: Class notes

José María Usategui

September 2023

Index

1- Preliminaries..3

Introduction. Real numbers. Absolute value and distance between real numbers. Intervals. Bounded and unbounded sets. Neighbourhoods. The Plane R^2. Implications. Mathematical proofs.

2- Algebra...10

The vector space R^n. Sum of vectors and product of a vector by a scalar. Product of vectors. Orthogonality. Norm and distance between two points. Linearly dependent/independent vectors. Basis. Matrices. Determinants. Minors and cofactors. Rank of a matrix. Systems of linear equations. Solutions of systems of linear equations. Inverse matrix. Definiteness of matrices.

3- Real-valued functions of a single variable..24

Real-valued functions of one real variable. Basic functions. Continuity. Concave and convex functions. Appendix: Limit and continuity.

4- Differential calculus with one variable..29

The derivative: Definition and meaning. The derivative: calculation. Linear approximation of a function. Second derivative and second order approximation of a function. Higher order derivatives and higher order approximations of a function. Differentiability and continuity. Mean-value Theorem. Global maximum and global minimum. Weierstrass' theorem. Local maximum and local minimum. Conditions for a local extreme.

5- Integral calculus with one variable..38

Indefinite integrals. Integration by parts. Integration by substitution or change of variable. Definite integrals. Introduction to differential equations.

6- Real-valued functions of several variables...42

Introduction. Continuity. Partial derivatives. Gradient. Hessian matrix. Differentiability. Total differential. The chain rule. Concavity and convexity of real-valued functions of several variables. Approximation of a function of several variables. Homogeneous functions. Euler's theorem. Multiple integrals. Appendix: Theorem of the implicit function and implicit differentiation.

7- Functions of R^n into R^m...50

The Jacobian matrix and the Jacobian determinant. Composition of functions.

1 Preliminaries

1.1 Introduction

Natural numbers: $\mathbf{N} = \{1, 2, 3, ...\}$

Integer numbers: $\mathbf{Z} = \mathbf{N} \cup \{0, -1, -2, -3, ...\} = \{0, 1, -1, 2, -2, 3, -3, ...\}$

Rational numbers: $\mathbf{Q} = \left\{\text{all fractions of the form} \frac{m}{n}, \text{ where } m \in \mathbf{Z} \text{ and } n \in \mathbf{N}\right\}$

Sum of rational numbers: $\frac{a}{b} + \frac{c}{d} = \frac{ad+cb}{bd}$ (warning: $\neq \frac{a+c}{b+d}$)

Product of rational numbers: $(\frac{a}{b})(\frac{c}{d}) = \frac{ac}{bd}$

Division of rational numbers: $(\frac{a}{b}) : (\frac{c}{d}) = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$

Order in $\mathbf{Q}$: $\frac{a}{b} > \frac{c}{d} \Leftrightarrow ad > cb$

Powers of rational numbers: a rational number x multiplied by itself n times ($n \in \mathbf{N}$) is written as x^n. If $x, y, r \in \mathbf{Q}$, $m, p \in \mathbf{Z}$ and $n, q \in \mathbf{N}$ we have:

i) If $y > 0$ then $y = x^n \Leftrightarrow x = y^{\frac{1}{n}} = \sqrt[n]{y}$,
ii) $(-x)^n \neq -x^n$ (example: $(-3)^2 \neq -3^2$),
iii) $x^{\frac{m}{n}} := (x^{\frac{1}{n}})^m$,
iv) $x^{\frac{m}{n}} x^{\frac{p}{q}} = x^{(\frac{m}{n})+(\frac{p}{q})} = x^{\frac{mq+pn}{nq}}$ (example: $3^{\frac{2}{3}} 3^{\frac{-3}{4}} = 3^{\frac{2}{3}-\frac{3}{4}} = 3^{\frac{-1}{12}} = \frac{1}{(3)^{\frac{1}{12}}}$),
v) $(x^{\frac{m}{n}})^{\frac{p}{q}} = x^{(\frac{m}{n})(\frac{p}{q})} = x^{\frac{mp}{nq}}$,
vi) $x^r y^r = (xy)^r$ and
vii) $(x + y)^r \neq x^r + y^r$.

But *powers* of rational numbers do *not* always are rational numbers! For instance, $2^{\frac{1}{2}} = \sqrt{2}$ is not a rational number. The *rational numbers* are *not enough* to express the exact measure of all magnitudes (another example: the quotient between the length of a circumference and its radius is not a rational number).

1.2 Real numbers

The *real numbers* $\mathbf{R}$ are required to express the exact measure of all magnitudes. *The real numbers are the result of adding the irrational numbers to the rational numbers.* The rational numbers may be defined as the set of decimal numbers with a finite number of decimals or with infinite decimals such that beyond a certain digit a finite sequence of digits repeats itself forever. The numbers with infinite decimals and such that there is not a finite sequence of digits that repeats itself forever beyond a certain digit are the *irrational numbers*.

Remark: there is *no* such thing like the number "next to" or "preceding to" any rational number! There are irrational numbers between any two consecutive rational numbers.

We have: $\mathbf{N} \subset \mathbf{Z} \subset \mathbf{Q} \subset \mathbf{R}$.

In the "real line" every point represents a real number.

1.3 Absolute value and distance between real numbers

The **absolute value of** $x \in \mathbf{R}$ is $|x| := \begin{cases} x, \text{ if } x > 0 \\ -x, \text{ if } x < 0 \end{cases}$

Properties of the absolute value:

i) For all $x \in \mathbf{R}$, $|x| \geq 0$.
ii) For all $x \in \mathbf{R}$, $|x| = |-x|$.
iii) For all $x, y \in \mathbf{R}$, $|x.y| = |x| \, . \, |y|$
iv) For all $x, y \in \mathbf{R}$, $|x + y| \leq |x| + |y|$ ($|x + y| = |x| + |y| \Leftrightarrow x$ and y are both positive or both negative)

The **distance** *between two real numbers*, $x, y \in \mathbf{R}$, denoted $d(x, y)$, is defined by $d(x, y) := |x - y|$.

Examples: $d(1, 6) = |1 - 6| = 5$, $d(6, 1) = |6 - 1| = 5$, $d(-1, 6) = |-1 - 6| = 7$, $d(-1, -6) = |-1 - (-6)| = 5$ and $d(6, -1) = |6 - (-1)| = 7$.

1.4 Intervals

For any $a, b \in \mathbf{R}$, the following subsets of $\mathbf{R}$ are called *intervals*:

- *Bounded intervals*:
  - $[a, b] := \{x \in \mathbf{R} : a \leq x \leq b\}$ (closed interval)
  - (a, b) or $]a, b[:= \{x \in \mathbf{R} : a < x < b\}$ (open interval)
  - $[a, b)$ or $[a, b[:= \{x \in \mathbf{R} : a \leq x < b\}$
  - $(a, b]$ or $]a, b] := \{x \in \mathbf{R} : a < x \leq b\}$

- *Unbounded intervals*:
  - $[a, \infty)$ or $[a, \infty[:= \{x \in \mathbf{R} : a \leq x\}$
  - $(-\infty, b]$ or $]-\infty, b] := \{x \in \mathbf{R} : x \leq b\}$
  - (a, ∞) or $]a, \infty[:= \{x \in \mathbf{R} : a < x\}$
  - $(-\infty, b)$ or $]-\infty, b[:= \{x \in \mathbf{R} : x < b\}$

1.5 Bounded and unbounded sets

A point/number $a \in \mathbf{R}$ is a *lower* (*upper*) *bound* of a subset $A \subset R$ if $x \geq a$ for all $x \in A$ ($x \leq a$ for all $x \in A$).

When a subset has a lower (upper) bound a any other number smaller (greater) than a is also a lower (upper) bound.

A set of real numbers $A \in \mathbf{R}$ is said to be *bounded below* (*above*) if there is a *lower* (*upper*) *bound* of A.

A set of real numbers $A \in \mathbf{R}$ is said to be *bounded* when it is *both* bounded bellow and bounded above.

The *maximum* (*minimum*) of a set $A \in \mathbf{R}$ is a number m such that: i) $m \in A$ and ii) $m \geq x$ ($m \leq x$) for all $x \in A$ (we write $\max A = m$ ($\min A = m$)).

Remark: Not all bounded above (below) set has a maximum (minimum); e.g., $\min\,[a, b) = a$, but $\nexists \max\,[a, b)$.

1.6 Neighbourhoods

A *neighbourhood* of a point/number $a \in \mathbf{R}$ of radious $r > 0$ is given by:

$$B_r(a) = \{x \in \mathbf{R} : |x - a| < r\}$$

In words: $B_r(a)$ is the set of points/numbers at a distance smaller than r from a.

Note that $B_r(a) = (a - r, a + r)$ or $]a - r, a + r[$ (open interval)

1.7 The Plane $\mathbf{R}^2$

The *plane*: $\mathbf{R}^2 = \{(x, y) : x, y \in \mathbf{R}\}$. Any point (x, y) may be placed in the plane. Any two-variable equation (i.e., the set of points (x, y) that satisfy it) can also be interpreted graphically over the plane.

Main curves in the plane R^2:

• **Straight line**: given two real numbers a and b, $y = ax + b$, where $x \in \mathbf{R}$, yields the straight line that crosses the y-axis at $(0, b)$ and has slope a.

- Example: $y = 0.5x + 2$ (i.e., the set $\{(x, y) \in \mathbf{R}^2 : y = 0.5x + 2\}$) yields the straight line that crosses the y-axis at $(0, 2)$ and has slope 0.5.

• **Circle**: given three real numbers a, b and r, $(x - a)^2 + (y - b)^2 = r^2$ yields a circle of radius r whose center is at (a, b).

• **Parabola**: given three real numbers a, b and c, $y = ax^2 + bx + c$ yields a parabola whose axis is paralell to the y-axis and $x = ay^2 + by + c$ yields a parabola whose axis is paralell to the x-axis.

1.8 Implications

Consider that P and Q are two properties, conditions, equations or propositions.

$P \Rightarrow Q$ means that if P occurs then Q is satisfied. In this case P is a *sufficient condition* for Q and Q is a *necessary condition* for P. Hence, P *may* occur only if Q is satisfied.

$P \Leftrightarrow Q$ means that if P is true then Q is true and viceversa (P and Q are equivalent)

• **Examples**:
i) $x = \sqrt{9} \Leftarrow x = 3$
ii) $x^2 < 4 \Rightarrow x < 2$
iii) $x(x^2 + 4) = 0 \Leftrightarrow x = 0$
iv) $x^2 > 0 \Leftarrow x > 0$

1.9 Mathematical proofs

To prove that $P \Rightarrow Q$ two methods can be used. In the *direct method* we assume P and try to prove Q. In the *indirect method* we proceed by absurd reduction: we assume that Q does not occur and try to prove that in that case P is not satisfied.

• **Example**: Prove that $x^2 - 3x + 7 < 0 \Rightarrow x > 0$ ($P : x^2 - 3x + 7 < 0$ and $Q : x > 0$)

Direct proof: $x^2 - 3x + 7 < 0 \Leftrightarrow 3x > x^2 + 7 \Rightarrow 3x > 0 \Leftrightarrow x > 0$

Indirect proof: If $x \leq 0 \Leftrightarrow 3x \leq 0 \Leftrightarrow -3x \geq 0 \Longrightarrow x^2 - 3x + 7 \geq 0$ (hence: no $Q \Rightarrow$ no P)

The method of *mathematical induction* is also used in some proofs. Consider that we want to prove that $P(n)$ is satisfied for all $n \in N$. We proceed in the following way:
i) prove $P(1)$
ii) assume that $P(k)$ is satisfied for $k \in N$
iii) prove that $P(k + 1)$ is satisfied

• **Example of mathematical induction**: Prove that $P(n) := 1 + 3 + 5 + ... + (2n - 1) = n^2$
i) for $n = 1$ it is $1 = 1^2 = 1$
ii) assume that $1 + 3 + 5 + ... + (2k - 1) = k^2$

iii) for $k+1$: $1+3+5+...+(2k-1)+(2(k+1)-1) = k^2+(2(k+1)-1) = k^2+2k+1 = (k+1)^2$

• **Another example**: prove by mathematical induction that $\sum_{j=1}^{n} j = \frac{n(n+1)}{2}$

2 Algebra

2.1 The vector space R^n

The n- dimensional vector space is:

$$R^n = \{(x_1, ..., x_n) : \text{ where all } x_i \text{ are real numbers}\}$$

The elements of R^n are called points or vectors and may be represented by $\mathbf{x} = (x_1, ..., x_n)$. The i^{th} component of $\mathbf{x}$ is x_i, with $i = 1, ..., n$.

If $n = 2$ or $n = 3$ then R^n allows a graphic interpretation.

Point $\mathbf{0} = (0, ..., 0) \in R^n$ with all components equal to 0 is fixed as the origin. Vector $\mathbf{x} \in R^n$ has associated an arrow that goes from $\mathbf{0}$ to $\mathbf{x}$.

A set in R^n is closed if and only if it includes its frontier. A set in R^n is bounded if it is bounded for each of its n components. It is defined that a closed and bounded set in R^n is a compact set.

2.2 Sum of vectors and product of a vector by a scalar

The **sum** of two vectors $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$ is:

$$\mathbf{x} + \mathbf{y} := (x_1 + y_1, x_2 + y_2, ..., x_n + y_n)$$

The following properties of the sum of vectors are immediate:

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$$
$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$$
$$\mathbf{x} + \mathbf{0} = \mathbf{x}$$

The **product by a scalar** λ (where $\lambda \in R$,i.e., λ is a real number) of a vector $\mathbf{x} = (x_1, ..., x_n)$ is:

$$\lambda\mathbf{x} := (\lambda x_1, ..., \lambda x_n)$$

The following properties of the product of a vector by a scalar are immediate:

$$1\mathbf{x} = \mathbf{x}$$
$$\lambda(\mathbf{x}+\mathbf{y}) = \boldsymbol{\lambda}\mathbf{x} + \lambda\mathbf{y}$$
$$(\lambda+\mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$$
$$(\lambda\mu)\mathbf{x} = \lambda(\mu\mathbf{x})$$
$$0\mathbf{x} = \mathbf{0}$$

Below $-\lambda\mathbf{x}$ is written to mean $(-\lambda)\mathbf{x} = (-\lambda x_1, ..., -\lambda x_n)$.

The sum of vectors and the product of a vector by a scalar have a clear interpretation. The sum of vectors $\mathbf{x}+\mathbf{y}$ corresponds to the vector represented by the diagonal of the parallelogram with sides $\mathbf{x}$ and $\mathbf{y}$. Moreover, $\lambda\mathbf{x}$ if $\lambda > 0$ ($\lambda < 0$) is a vector in the straight line through $\mathbf{0}$ and $\mathbf{x}$, at a distance from $\mathbf{0}$ equal to λ-times the distance from $\mathbf{x}$ to $\mathbf{0}$ in the same (opposite) direction as $\mathbf{x}$.

2.3 Product of vectors. Orthogonality.

The **scalar or inner product of two vectors** $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$ is:

$$\mathbf{x}.\mathbf{y} := x_1y_1 + x_2y_2 + ... + x_ny_n$$

The following properties of the inner product of vectors are immediate:

$$\mathbf{x}.\mathbf{y} = \mathbf{y}.\mathbf{x}$$
$$\mathbf{x}.(\mathbf{y}+\mathbf{z}) = \mathbf{x}.\mathbf{y} + \mathbf{x}.\mathbf{z}$$
$$(\lambda\mathbf{x})\mathbf{y} = \boldsymbol{\lambda}(\mathbf{x}.\mathbf{y})$$

Two **vectors x** and **y** in R^n are **orthogonal** if $\mathbf{x}.\mathbf{y} = \mathbf{0}$. When $\mathbf{x}.\mathbf{y} = \mathbf{0}$ in R^2 or in R^3 vectors $\mathbf{x}$ and $\mathbf{y}$ are **perpendicular**.

This gives us another way of specifying a straight line or a plane. If $\mathbf{a}$ and $\mathbf{s}$ are vectors in R^2, the straight line that crosses point $\mathbf{a}$ and is orthogonal to $\mathbf{s}$ is the set of points $\mathbf{x}$ that satisfy the equation:

$$\mathbf{s}.(\mathbf{x}-\mathbf{a}) = 0,$$

or, equivalently, $\mathbf{s}.\mathbf{x} = \mathbf{s}.\mathbf{a}$. If $\mathbf{a}$ and $\mathbf{s}$ are vectors in R^3 those equations would give the equation of a plane crossing $\mathbf{a}$ and orthogonal to $\mathbf{s}$.

Example 1: The equation of the plane that crosses point $\mathbf{a} = (3, 1, -1)$ and is orthogonal to $\mathbf{s} = (1, 2, 3)$ is given by $\mathbf{s}.\mathbf{x} = \mathbf{s}.\mathbf{a}$, that is:

$$(1, 2, 3).(x_1, x_2, x_3) = (1, 2, 3).(3, 1, -1),$$

or, equivalently, $x_1+2x_2+3x_3 = 2$. Note that the coefficients in this equation of the plane are the components of the vector orthogonal to that plane (the same happens for a straight line in R^2).

2.4 Norm and distance between two points

The **norm of a vector** $\mathbf{x} \in R^n$, denoted by $\|\mathbf{x}\|$, is the number:

$$\|\mathbf{x}\| := \sqrt{(x_1)^2 + (x_2)^2 + ... + (x_n)^2}$$

This number represents the distance from point $\mathbf{x}$ to the origin or the length of vector $\mathbf{x}$.

The **distance between two points** $\mathbf{x} \in R^n$ and $\mathbf{y} \in R^n$ is given by:

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + ... + (x_n - y_n)^2}$$

A *neighbourhood* of a point $\mathbf{x} \in R^n$ of radious $r > 0$ is given by:

$$B_r(\mathbf{x}) = \{\mathbf{y} \in R^n : d(\mathbf{x}, \mathbf{y}) < r\}$$

2.5 Linearly dependent/independent vectors. Basis.

Any vector that can be obtained combining sum and product by scalars of other vectors is called a "**linear combination**" of the initial ones. The following **definition** can be stated: A vector $\mathbf{x} \in R^n$ is a *linear combination* *of m vectors* $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m \in R^n$ if there are m real numbers $\lambda_1, \lambda_2, ..., \lambda_m \in R$ such that:

$$\mathbf{x} = \lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2 + ... + \lambda_m \mathbf{a}_m$$

Vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m \in R^n$ are said to be **linearly dependent (independent)** if one (none) of them is a linear combination of the others. The following formulation of this notion is equivalent: Vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m \in R^n$ are *linearly independent* if:

$$\lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2 + ... + \lambda_m \mathbf{a}_m = \mathbf{0} \text{ only if } \lambda_1 = \lambda_2 = ... = \lambda_m = 0$$

Hence, vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m \in R^n$ are *linearly dependent* when $\lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2 + ... + \lambda_m \mathbf{a}_m = \mathbf{0}$ for some coefficients λ_i not all of them 0.

A **basis** in R^n is a set of n vectors *linearly independent* and such that any other vector is dependent on them. In other terms, a basis is a set of vectors such that: (i) any other vector is a linear combination of them and (ii) none of them is a linear combination of the others. Hence: (i) starting from that set of vectors any vector is reachable by combining product by scalars and sum, but (ii) this is not possible any longer if one vector is eliminated from the set (for instance, the one eliminated is not reachable from the other vectors in the basis).

Example 2: Vectors $(1, 0)$ and $(0, 1)$ form a basis of R^2; but there are infinite many other basis. For instance, $(1, 1)$ and $(-1, 1)$ form another basis.

We will later see how to know whether a set of vectors in R^n form a basis or not.

2.6 Matrices

An m x n *matrix* **A** is a set of real numbers (the *elements* or *entries*) arranged in a rectangle of m *rows* and n *columns*, that is:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & ... & a_{mn} \end{pmatrix}$$

The set of all matrices of *dimension* m x n is denoted $M(m,n)$. For matrix **A** above the notation $\mathbf{A} = (a_{ij})_{\substack{i=1,2,...,m \\ j=1,2,...,n}}$ is also used, or just $\mathbf{A} = (a_{ij})$ when the dimension is clear from the context. When $m = n$ the matrix is said to be a *square matrix*. A square **matrix** is **symmetric** if and only if $a_{ij} = a_{ji}$, whenever $i \neq j$.

The *main diagonal* of a square matrix consists of the elements a_{ii}, $i = 1, 2, ..., n$. A square matrix is a *diagonal matrix* if $a_{ij} = 0$, whenever $i \neq j$, and the n x n *identity square matrix*, denoted $\mathbf{I}_n$, is the diagonal matrix such that $a_{ii} = 1$, $i = 1, 2, ..., n$, that is:

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & ... & 0 \\ 0 & 1 & ... & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & 1 \end{pmatrix}$$

The following **operations** are defined **with matrices**:

- The **sum or addition** of two matrices $\mathbf{A}, \mathbf{B} \in M(m,n)$ is the matrix:

$$\mathbf{A} + \mathbf{B} = (a_{ij}) + (b_{ij}) := (a_{ij} + b_{ij})$$

- The **product of a matrix by a scalar** (i.e., a real number) λ is the matrix:

$$\lambda \mathbf{A} = \lambda (a_{ij}) := (\lambda a_{ij})$$

Example 3: $3 \begin{pmatrix} 1 & 0 & -2 \\ 2 & 3 & 0 \end{pmatrix} - 2 \begin{pmatrix} -2 & 4 & 2 \\ 5 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 7 & -8 & -10 \\ -4 & 9 & -2 \end{pmatrix}$.

- The **product of two matrices** $\mathbf{A} \in M(m,n)$ and $\mathbf{B} \in M(n,r)$, denoted **A**.**B** (the dot "."is often omitted, just writing **AB**) is the m x r matrix:

$$\mathbf{A}.\mathbf{B} = (a_{ij}).(b_{ij}) := (c_{ij})$$

where $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + ... + a_{in}b_{nj}$. That is to say, c_{ij} is the element that results from multiplying row i of $\mathbf{A}$ and column j of $\mathbf{B}$:

$$\begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{pmatrix} \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{pmatrix}$$

Note that for the product to make sense the number of columns of the first matrix has to coincide with the number of rows of the second.

Example 4: $\begin{pmatrix} 1 & 0 & -2 \\ 2 & 3 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 & 2 \\ 2 & 1 & 0 & -2 \\ -1 & -2 & 4 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 4 & -9 & -4 \\ 8 & 3 & -2 & -2 \end{pmatrix}$.

The following **properties** are easy to check:

- Assuming that all matrices are of the same dimension m x n, for the *addition of matrices* we have:
(i) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
(ii) $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
(iii) $\mathbf{A} + \mathbf{O} = \mathbf{A}$ (where $\mathbf{O}$ denotes the $m\ x\ n$ 0-matrix whose entries are all 0).

- Assuming that all matrices are of the same dimension m x n, for the *product by a scalar* we have:
(i) $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B}$
(ii) $(\lambda + \mu)\mathbf{A} = v\mathbf{A} + \mu\mathbf{A}$
(iii) $\lambda(\mu\mathbf{A}) = (\lambda\mu)\mathbf{A}$

- Assuming that all matrices are of the required dimension for the product to make sense, we have:
(i) $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
(ii) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
(iii) $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$
(iv) $\mathbf{AI}_n = \mathbf{I}_n\mathbf{A} = \mathbf{A}$

- Note that in general $\mathbf{AB} \neq \mathbf{BA}$, as the following example shows:

Example 5: Let $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}$. Note that:

$$\mathbf{AB} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 1 \\ 3 & 0 \end{pmatrix}$$
$$\mathbf{BA} = \begin{pmatrix} 3 & 0 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 6 & 3 \\ 5 & 2 \end{pmatrix}$$

- Note that $\mathbf{AB} = \mathbf{O} \nRightarrow \mathbf{A} = \mathbf{O}$ or $\mathbf{B} = \mathbf{O}$, as the following example shows:

Example 6: Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} -2 & 0 \\ 1 & 0 \end{pmatrix}$. It follows that:

$$\mathbf{AB} = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} -2 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

The **transpose of** an m x n **matrix** $\mathbf{A} = (a_{ij})_{\substack{i=1,2,\ldots,m \\ j=1,2,\ldots,n}}$ is the n x m matrix $\mathbf{A}^T = (a_{ji})_{\substack{j=1,2,\ldots,n \\ i=1,2,\ldots,m}}$. That is to say, $\mathbf{A}^T$ is the result of taking the first row of $\mathbf{A}$ as the first column of $\mathbf{A}^T$, the second row as the second column, etc. As properties of the transpose we have:

(i) $(\mathbf{A}^T)^T = \mathbf{A}$

(ii) $(\mathbf{A}.\mathbf{B})^T = \mathbf{B}^T.\mathbf{A}^T$ (note the change of order!)

(iii) $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$

(iv) $(\lambda\mathbf{A})^T = \lambda\mathbf{A}^T$

(v) A is symmetric $\Leftrightarrow \mathbf{A} = \mathbf{A}^T$

2.7 Determinants. Minors and cofactors.

In this section we deal with square matrices. The **determinant of the matrix** $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ is denoted by:

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} := a_{11}a_{22} - a_{21}a_{12}.$$

The following are **properties of the determinants of** 2 **x** 2 **matrices** that are very easy to check. Let A be a 2 x 2 matrix:

- (i) If $\mathbf{A}'$ is the result of multiplying a row or a column of $\mathbf{A}$ by a real number λ we have $|\mathbf{A}'| = \lambda\,|\mathbf{A}|$. For instance:

$$\begin{vmatrix} 3(2) & -2 \\ 3(4) & 1 \end{vmatrix} = 30 = 3\begin{vmatrix} 2 & -2 \\ 4 & 1 \end{vmatrix}$$

- (ii) If $\mathbf{A}'$ is the result of adding an arbitrary row-vector (or an arbitrary column-vector) to a row (or a column) of $\mathbf{A}$ we have:

$$|\mathbf{A}'| = |\mathbf{A}| + |\mathbf{A}''|$$

where $\mathbf{A}''$ is the matrix that results by replacing that row (or column) by the one added. For instance:

$$\begin{vmatrix} 2+6 & -2 \\ 4+5 & 1 \end{vmatrix} = \begin{vmatrix} 2 & -2 \\ 4 & 1 \end{vmatrix} + \begin{vmatrix} 6 & -2 \\ 5 & 1 \end{vmatrix}$$

- (iii) If $\mathbf{A}'$ is the result of interchanging the two rows or the two columns of $\mathbf{A}$ we have:

$$|\mathbf{A}'| = (-1)\,|\mathbf{A}|\,.$$

For instance:

$$\begin{vmatrix} 2 & -2 \\ 4 & 1 \end{vmatrix} = -\begin{vmatrix} -2 & 2 \\ 1 & 4 \end{vmatrix}$$

- (iv) From the definition it obviously follows:

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

These four simple properties may be extended to n x n matrices and they characterize univocally the determinant each matrix.

From the four properties (i) to (iv) (see above) *other basic properties of determinants* can be derived, among which we underline the following:

- (v) If an n x n matrix has two rows or two columns identical (or proportional) its determinant is 0 (this follows from property (iii)). Hence, by considering also (i) and (ii) it follows that if a multiple of one row (of one column) is added to a different row (to a different column) of $\mathbf{A}$ then the determinant of the matrix obtained is equal to $|\mathbf{A}|$.

- (vi) The determinant of the product of two square n x n matrices **A** and B is the product of their determinants:

$$|\mathbf{A}.\mathbf{B}| = |\mathbf{A}|\,|\mathbf{B}|\,.$$

- (vii) The determinant of an n x n matrix A and that of its transpose coincide:

$$\left|\mathbf{A}^T\right| = |\mathbf{A}|$$

From conditions (i), (ii) and (v) it is obtained:

Proposition 1: *If the determinant of a matrix is different from* 0 *(equal to* 0*) then the column vectors of that matrix are independent (dependent) and the same follows for the row vectors.*

Remark 1: *As a consequence of Proposition 1 a set of n vectors is a basis of R^n if and only if the determinant of the matrix that has those vectors as columns (or as rows) is different from* 0.

Let **A** be a square n x n matrix:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

The **minor** $|\mathbf{A}_{ij}|$ is the *determinant of the* $(n-1)$ x $(n-1)$ *matrix* that results by eliminating row i and column j in **A**. The **cofactor** C_{ij} is the number:

$$C_{ij} := (-1)^{i+j}\,|A_{ij}|\,.$$

The following definition is useful for some analyses:

Definition: The **leading principal minors** of a square n x n matrix $\mathbf{A} = (a_{ij})_{\substack{i=1,2,\ldots,m\\ j=1,2,\ldots,n}}$ are $|a_{11}|$, $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, $\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$,..., and $\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$.[1]

[1] The determinant of a submatrix of order r x r of a square n x n matrix A obtained

2.8 Rank of a matrix

Let **A** be an m x n matrix. A **minor of order** r ($1 \leq r \leq n$) of **A** is the *determinant* of an r x r square matrix obtained from **A** by deleting any $m - r$ rows and any $n - r$ columns.

The **rank of a matrix A** ($rank(\mathbf{A})$) is the maximal number of linearly independent columns in **A**. It coincides with the maximal number of linearly independent rows in **A**. The following result is obtained:

Proposition 2: *Let A be an m x n matrix, the* **rank of A** *is the* ***highest order of a non null minor***.

2.9 Systems of linear equations

A **system of** m **linear equations with** n **variables or unknowns**, x_i ($i = 1, ..., n$), is a system of the form:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n = b_m \end{cases} \tag{1}$$

where the a_{ij} (the *coefficients*) and b_i (the *right-hand sides*) are given real numbers. A **solution of the system** is any point or vector $(x_1, x_2, ..., x_n)$ that satisfies all the equations. In some cases there is no solution. If there is **no solution** then the **system** is said to be **inconsistent**. When **at least a solution does exist** the **system** is said to be **consistent**. In this case, when there is a **unique solution** the system is said to be **determinate** and otherwise (i.e., **when there are infinite solutions**) the system is said to be **indeterminate**.

by eliminating $n - r$ rows and $n - r$ columns is called a principal minor of order r of A if the labels of the rows and columns eliminated coincide (for instance, if $n = 5$ and $r = 3$ rows 2 and 4 and columns 2 and 4 are eliminated to obtain the 3 x 3 submatrix). If $r < n$ there are several principal minors of order r of matrix A.

A linear **system** (1) is said to be **homogeneous** if all right hand sides are 0, that is, $b_i = 0$ for all $i = 1, ..., m$. In an homogeneous system, there always exists at least one solution: $x_1 = x_2 = ... = x_n = 0$, called the *trivial solution.* Therefore consistency is guaranteed in that system.

In a system as (1) the *matrix of coeficients* is the matrix:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

and we call the *augmented matrix* to the matrix that results by adding a column consisting of the right hand sides of the equations (that we separate with a line):

$$(\mathbf{A}\,|\mathbf{b}) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right),$$

where **b** denotes the m x 1 column-vector of the right-hand sides:

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

If similarly we denote by **x** the n x 1 column-vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

using the product of matrices introduced in Section 2.6 the system (1) can be written in the following form:

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \tag{2}$$

that is, $\mathbf{Ax} = \mathbf{b}$. Alternatively the system can be expressed in this way:

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + ... + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{3}$$

that is, if $\mathbf{a}_j$ denotes the j column-vector in $\mathbf{A}$,

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + ... + x_n\mathbf{a}_n = \mathbf{b} \tag{4}$$

Evidently, (1), (2) and (3) or (4) are equivalent forms of expressing the same system.

Remark 2: *The equivalent expressions (3) and (4) provide another interpretation of a linear system: Saying that system (1) is consistent is equivalent to saying that vector* $\mathbf{b}$ *is a linear combination of* $\mathbf{a}_1$, $\mathbf{a}_2$,..., $\mathbf{a}_n$.

2.10 Solutions of systems of linear equations

There are three possible situations when facing the resolution of a system of linear equations with n unknowns:

- Case 1: $rank(\mathbf{A}) = rank(\mathbf{A}\,|\mathbf{b}) = n$. Then *the system is consistent and determinate*: there exists a unique solution.

- Case 2: $rank(\mathbf{A}) = rank(\mathbf{A}\,|\mathbf{b}) < n$. Then *the system is consistent and indeterminate*: there exist infinite solutions and there are $n - rank(\mathbf{A})$ degrees of freedom.

- Case 3: $rank(\mathbf{A}) < rank(\mathbf{A}\,|\mathbf{b})$. Then *the system is inconsistent*: no solution exists.

In the particular case of *linear homogeneous systems* (i.e., when $\mathbf{b} = \mathbf{0}$) we have:

- Case H1: $rank(\mathbf{A}) = n$. Then *there only exists the trivial solution* $\mathbf{x} = \mathbf{0}$.

- Case H2: $rank(\mathbf{A}) < n$. *Then there exist infinite solutions and there are* $n - rank(\mathbf{A})$ *degrees of freedom.*

In case H1 column-vectors $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$, are linearly independent, while those vectors are linearly dependent in case H2.

The following result is obtained:

Proposition 3: For a n x n square matrix:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & ... & a_{nn} \end{pmatrix}$$

$rank(\mathbf{A}) = n$ is a *necessary and sufficient condition* for each of the following situations:

(i) The system $\mathbf{Ax} = \mathbf{b}$ has a unique solution whatever the right-hand sides column $\mathbf{b}$.

(ii) The column vectors of $\mathbf{A}$, $\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n$, form a basis of R^n (i.e., they are linearly independent and any other vector is a linear combination of them).

(iii) $|\mathbf{A}| \neq 0$.

2.11 Inverse matrix

The **inverse of a square** n x n **matrix A** is a matrix, denoted by $\mathbf{A}^{-1}$ (when it exists!), such that $\mathbf{A}^{-1}.\mathbf{A} = \mathbf{A}.\mathbf{A}^{-1} = \mathbf{I}_n$. Matrix $\mathbf{A}^{-1}$ may be obtained as follows:

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{C_{11}}{|A|} & \frac{C_{21}}{|A|} & ... & \frac{C_{n1}}{|A|} \\ \frac{C_{12}}{|A|} & \frac{C_{22}}{|A|} & ... & \frac{C_{n2}}{|A|} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{C_{1n}}{|A|} & \frac{C_{2n}}{|A|} & ... & \frac{C_{nn}}{|A|} \end{pmatrix}$$

where C_{ij} is the ij-cofactor. Thus A has an inverse matrix when $|\mathbf{A}| \neq 0$ (hence, according to Proposition 3, A has an inverse matrix if and only if $rank(\mathbf{A}) = n$).

The inverse of $\mathbf{A}$ can be obtained in three steps: i) Obtain the transpose $\mathbf{A}^T$, ii) replace each entry in $\mathbf{A}^T$ by the corresponding cofactor, and iii) divide all entries by $|\mathbf{A}|$.

The calculation of the inverse provides a different way of solving a linear system of n equations with n unknowns when the system is consistent and determinate, since:

$$\mathbf{Ax} = \mathbf{b} \Rightarrow \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{b} \Rightarrow \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}.$$

The following are properties relatives to the inverse and how it interacts with other operations:
(i) $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
(ii) $(\mathbf{A}.\mathbf{B})^{-1} = \mathbf{B}^{-1}.\mathbf{A}^{-1}$ (note the change of order!)
(iii) $(\mathbf{A}^{T})^{-1} = (\mathbf{A}^{-1})^{T}$
(iv) $(\lambda\mathbf{A})^{-1} = \frac{\mathbf{A}^{-1}}{\lambda}$, for any $\lambda \in R$
(v) if $\mathbf{A}$ is symmetric then $\mathbf{A}^{-1}$ is also symmetric

2.12 Definiteness of matrices

An n x n **matrix A** is:
(i) negative semidefinite if $x^T Ax \leq 0$ for all $x \in R^n$ and there is $\tilde{x} \in R^n$ such that $\tilde{x} \neq 0$ and $\tilde{x}^T A\tilde{x} = 0$,
(ii) negative definite if $x^T Ax < 0$ for all $x \in R^n$ and $x \neq 0$,
(iii) positive semidefinite if $x^T Ax \geq 0$ for all $x \in R^n$ and there is $\tilde{x} \in R^n$ such that $\tilde{x} \neq 0$ and $\tilde{x}^T A\tilde{x} = 0$, and
(iv) positive definite if $x^T Ax > 0$ for all $x \in R^n$ and $x \neq 0$.

3 Real-valued functions of a single variable

3.1 Real-valued functions of one real variable

Definition: *A magnitude or variable y is said to be a* ***function*** *of another x if to each value of x corresponds a single value of y.*

When y is a function of x we write $y = f(x)$ and say that "*f maps x into $f(x)$*".

In Chapters 3, 4 and 5 we deal with *functions where both variables x and y take real numerical values*, i.e., $x, y \in \mathbf{R}$. These functions are called "*real-valued functions of one real variable*".

When $D \subset \mathbf{R}$ is the set of possible values of x we write $f : D \to \mathbf{R}$, to express that $x \in D \subset \mathbf{R}$ and $y = f(x) \in \mathbf{R}$. The set D is called the **"domain" of** f.

If $A \subset D$ then $f(A) = \{f(x) : x \in A\}$. The set $f(A)$ is called the "**image**" *of A by f*. In particular, $f(D)$ is called the **"range" of** f.

Given a real-valued function $y = f(x)$ or $f : D \to \mathbf{R}$, the **"graph" of the function** f is the subset of $\mathbf{R}^2$:

$$\textit{graph}\ (f) := \left\{(x, y) \in \mathbf{R}^2 : x \in D,\ y = f(x)\right\} = \{(x, f(x)) : x \in D\}$$

(hence, the graph of a function can be represented graphically by the corresponding set of points in the plane).

• **Examples**:

i) Find the domain and range of $y = f(x) = \frac{1}{x+4}$.

Domain: As the function is not defined for $x = -4$ it follows that $D = \{x \in \mathbf{R} : x \neq -4\}$

Range: Since $y = \frac{1}{x+4} \Leftrightarrow x + 4 = \frac{1}{y}$ it follows that y can never be 0. Hence, $f(D) = \{y \in \mathbf{R} : y \neq 0\}$.

ii) Find the domain and range of $y = f(x) = \sqrt{3x+9}$.

Domain: As the function is not defined for $3x + 9 < 0$ and as $3x + 9 < 0 \Leftrightarrow x < -3$, it follows that $D = \{x \in \mathbf{R} : x \geq -3\}$

Range: As $y = \sqrt{3x+9}$ and $x \geq -3$ it follows that $f(D) = \{y \in \mathbf{R}\}$

3.2 Basic functions

• **Linear functions**: A linear function is one of the form $y = ax + b$ $(a \neq 0)$. Their graphs are *straight lines*.

• **Quadratic functions**: A quadratic function is one of the form $y = ax^2 + bx + c$ $(a \neq 0)$. Their graphs are *parabolas*.

• **Polynomials**: A polynomial of degree n is a function of the form $y = p(x) = a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + ... + a_2x^2 + a_1x^1 + a_0$ $(a_n \neq 0)$. The following results hold for polynomials:

- **Result 1**: The polynomial $p(x)$ *has the factor* $(x - \lambda)$ if and only if $p(\lambda) = 0$.

Hence, a polynomial of degree n has at most n different factors of the form $(x - \lambda_i)$, with $i = 1, 2, ..., n$.

- **Result 2**: A polynomial of degree n *is defined by* the numerical value of $p(x)$ for some particular x and n different factors λ_i of the form $(x - \lambda_i)$.

- **Result 3**: A polynomial of degree n *is defined by* $n+1$ points $(x, p(x))$ in R^2 that are crossed by the graph of $p(x)$.

• **Rational functions**: A rational function is a function $y = f(x)$ where $f(x)$ is a *quotient of polynomials*. Example: $y = \frac{3x^2-5x+1}{x^3-4x}$.

• **Power functions**: A power function is one of the form $y = cx^a$ where $a, c \in \mathbf{R}$ are given. When $a \in N$, x can take any real value, *otherwise* $x > 0$ *is required.*

• **Exponential functions**: An exponential function is a function of the form $y = ae^x$ where $a \in \mathbf{R}$ is given.

• **Logarithmic functions**: A logarithmic function is a function of the form $y = a \log x$, with $x > 0$, where $a \in \mathbf{R}$ is given.

• **Trigonometric functions**: The trigonometric functions are $y = \sin x$, $y = \cos x$ and $y = tgx = \frac{\sin x}{\cos x}$.

3.3 Continuity

• The basic idea (roughly speaking): A *function* $y = f(x)$ is said to be *continuous* if as x varies in a continuous way (i.e., without brusque leaps) $f(x)$ varies also in a continuous way, or if *"its graph is not broken"*.

• **Definition**: *A* ***function*** $f(x)$ *is* ***continuous at a point*** x_0 *if for all neighbourhood* V *of* $f(x_0)$ *there is a neighbourhood* B *of* x_0 *such that* $f(B) \subset V$.

Hence (roughly speaking again): a function is continuous at a point x_0 if "it maps points close to x_0 (i.e., in the vicinity of x_0) into points close to $f(x_0)$ (i.e., in the vicinity of $f(x_0)$)". The notion of neighbourhood, i.e., any interval of the form $B_r(x_0) = (x_0 - r, x_0 + r)$, captures the idea of "vicinity".

• A function may be continuous at a point but not at another. A **function** is said to be **continuous on a set** when it is continuous at any point in that set. A **function** is said to be **continuous** when it is continuous on its domain.

• Most of functions we will be dealing with are continuous because:

- i) *linear and quadratic functions, polynomials, rational functions (where the denominator is* $\neq 0$*), power functions, exponential functions, logarithmic functions and trigonometric functions are continuous in their domains*,

- ii) *the sum, substraction, product and quotient (where the denominator is* $\neq 0$*) of continuous functions are continuous*,

- iii) *the power of a continuous functions are continuous: if* f *is continuous then* $c(f(x))^a$ *is continuous*, and

- iv) *the composition of continuous functions is a continuous function* (*Definition*: If $z = g(y)$ and $y = f(x)$ are two functions such that the range of f is contained in the domain of g, then we can define the *composition of* f *and* g, denoted by $g \circ f$ as $z(x) = (g \circ f)(x) := g(f(x))$).

Combining i)-iv) the continuity of many functions may be checked.

• *The inverse of a continuous function may not be a function* (if $y = f(x)$ the inverse of f, denoted f^{-1}, is $x = f^{-1}(y)$, but f^{-1} may not be a function even if f is continuous→ Example in class).

• Two nice properties of a continuous function are the following:

Bolzano's Theorem: If a function $y = f(x)$ is *continuous* on an interval $[a, b]$ and $f(a)$ and $f(b)$ *have different sign*, then there is at least one point c ($a < c < b$) where $f(c) = 0$.

Intermediate value's Theorem: If a function $y = f(x)$ is *continuous* on an interval $[a, b]$ then for each number γ *(strictly) between* $f(a)$ and $f(b)$ there exist at least one point c ($a < c < b$) where $f(c) = \gamma$.

3.4 Concave and convex functions

• Remember that a **set** D in **R** is **convex** if and only if for any $a, b \in D$ and for every scalar $\theta \in [0, 1]$ it is true that $\theta a + (1 - \theta)b \in D$. Then we have:

Definition: *Let* f *be a real-valued function defined on a convex set* D *in* R*. The function* f *is called a:*

i) ***concave function*** *if, for all* $a, b \in D$ *and* $0 \leq \theta \leq 1$, $f\left[\theta a+(1-\theta)b\right] \geq \theta f(a)+(1-\theta)f(b)$

ii) ***strictly concave function*** *if, for all* $a, b \in D$ *and* $0 < \theta < 1$, $f\left[\theta a+(1-\theta)b\right] > \theta f(a)+(1-\theta)f(b)$

iii) ***convex function*** *if, for all* $a, b \in D$ *and* $0 \leq \theta \leq 1$, $f\left[\theta a+(1-\theta)b\right] \leq \theta f(a)+(1-\theta)f(b)$

iv) ***strictly convex function*** *if, for all* $a, b \in D$ *and* $0 < \theta < 1$,

$$f\left[\theta a+(1-\theta)b\right] < \theta f(a)+(1-\theta)f(b)$$

3.5 Appendix: Limit and continuity

• In general, we have the following:

Definition: *We say that the* ***limit of a function*** $f(x)$ ***at a point*** x_0 *(or when "x tends to* x_0*") is* L*, and write* $\lim_{x\to x_0} f(x) = L$*, if for all neighbourhood* B *of* L *there is a neighbourhood* A *of* x_0 *such that* $f(A-\{x_0\}) \subset B$.

This allows for an *alternative definition of continuity*:

Definition: *A* ***function*** $f(x)$ *is* ***continuous at a point*** x_0 *of its domain if: i) there exists* $\lim_{x\to x_0} f(x)$*, and ii)* $\lim_{x\to x_0} f(x) = f(x_0)$.

• If $\lim_{x\to x_0} f(x) = L_1$ and $\lim_{x\to x_0} g(x) = L_2$ then:

i) $\lim_{x\to x_0}(f(x)+g(x)) = L_1 + L_2$,

ii) $\lim_{x\to x_0}(f(x)g(x)) = L_1 L_2$, and

iii) $\lim_{x\to x_0}(\frac{f(x)}{g(x)}) = \frac{L_1}{L_2}$.

4 Differential calculus with one variable

4.1 The derivative: definition and meaning

Given a function $y = f(x)$, the rate (or speed) of change of y with respect to x is given by the derivative, namely:

Definition: *The* ***derivative*** *of a function* $y = f(x)$ *(or "the derivative of* y *with respect to* x*")* ***at a point*** x_0 *is the value of the limit* (if it exists!):

$$f'(x_0) = \frac{dy}{dx}(x_0) := \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

When this limit does exist we say that $f(x)$ is *derivable at* x_0; and when a function is derivable at any point within a certain region A we say that it is *derivable on* A.

Given a function $y = f(x)$ derivable at a point x_0:

• $f'(x_0)$ represents the rate of change of y with respect to x at x_0. Therefore the derivative informs about the impact on $f(x)$ of changes in x. If $f'(x_0) > 0$ a slight *increase* of x entails a slight *increase* of $y = f(x)$, and a slight decrease of x entails a slight decrease of $y = f(x)$. If $f'(x_0) < 0$ a slight *increase* of x entails a slight *decrease* of $y = f(x)$, and a slight decrease of x entails a slight increase of $y = f(x)$.The magnitude of $f'(x_0)$ tells about the magnitud of that impact.

• The derivability of $y = f(x)$ at x_0 means that the graph of f is smooth enough as to have a tangent at $(x_0, f(x_0))$.

• $f'(x_0)$ is the slope of the tangent to the the graph of f at $(x_0, f(x_0))$.

4.2 The derivative: calculation

The *calculation of derivatives* is based on *two elements*: (i) A "table of derivatives" consisting of well-known derivatives, and (ii) The rules to derive a sum, a product or a quotient of functions.

Table of derivatives (I):

$$\frac{d}{dx}x^n = nx^{n-1}.$$
$$\frac{d}{dx}\ln x = \frac{1}{x},$$
$$\frac{d}{dx}e^x = e^x,$$
$$\frac{d}{dx}a^x = a^x \ln a,$$
$$\frac{d}{dx}\sin x = \cos x,$$
$$\frac{d}{dx}\cos x = -\sin x.$$

Rules to derive the sum, product and quotient of functions:

Proposition 1: If $f(x)$ and $g(x)$ are derivable functions, then $f(x)+g(x)$, $f(x).g(x)$ and $f(x)/g(x)$ (whenever $g(x) \neq 0$) *are also derivable* and:

$$\frac{d}{dx}(f(x)+g(x)) = f'(x)+g'(x)$$

$$\frac{d}{dx}(f(x).g(x)) = f'(x).g(x)+g'(x).f(x)$$

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x).g(x)-g'(x).f(x)}{(g(x))^2}$$

The chain rule and a new Table of derivatives:

Proposition 2 ("Chain rule"): The composition z of two derivable functions, f and g, is also derivable and we have:

$$\frac{dz(x)}{dx} = (g \circ f)'(x) = \frac{d}{dx}(g(f(x)) = g'(f(x)).f'(x)$$

Table of derivatives (II): By combining Table I of derivatives and the chain rule the following table of derivatives can be obtained:

$$\frac{d}{dx}f(x)^n = nf(x)^{n-1}f'(x),$$

$$\frac{d}{dx}(\ln f(x)) = \frac{f'(x)}{f(x)},$$

$$\frac{d}{dx}(e^{f(x)}) = f'(x)e^{f(x)},$$

$$\frac{d}{dx}(a^{f(x)}) = f'(x)a^{f(x)}\ln a,$$

$$\frac{d}{dx}(\sin f(x)) = f'(x)\cos f(x),$$

$$\frac{d}{dx}(\cos f(x)) = -f'(x)\sin f(x).$$

Logarithmic derivation:

The problem: How to derive $F(x) = f(x)^{g(x)}$? (e.g. $F(x) = (x^2 - 2x)^{2x}$)

- First, *take logarithms*: $\ln F(x) = \ln(f(x)^{g(x)}) = g(x)\ln f(x)$
- Second, *derive* both sides:

$$\frac{F'(x)}{F(x)} = g'(x)\ln f(x) + g(x)\frac{f'(x)}{f(x)}$$

- Therefore:

$$F'(x) = F(x)\left[g'(x)\ln f(x) + g(x)\frac{f'(x)}{f(x)}\right] = f(x)^{g(x)}\left[g'(x)\ln f(x) + g(x)\frac{f'(x)}{f(x)}\right]$$

4.3 Differentiability and continuity

Proposition 3: *If f is derivable (differentiable) at x_0 then f is continuous at x_0.*

Result: A continuous function may not be derivable.

Example: The function $f(x) = |x - 3| + 1$ is continuous at $x = 3$ but it is not derivable at that point. The derivative does not exist at $x = 3$ because the function has a kink at that point (the graph is not smooth and there is not a tangent at that point).

Continuity is a necessary condition for differentiability but continuity is not a sufficient condition for differentiability. We have: $f(x)$ is derivable at $x_0 \Rightarrow f(x)$ is continuous at x_0. However, $f(x)$ is continuous at $x_0 \not\Rightarrow f(x)$ is derivable at x_0.

4.4 Mean-value Theorem

Let f be a differentiable function in $[a, b]$. If $a < b$ there exists θ such that $a < \theta < b$ and:

$$f'(\theta) = \frac{f(b) - f(a)}{b - a}$$

4.5 Linear approximation of a function

As we saw: *The derivability of $f(x)$ at x_0 means that the graph of f is smooth enough as to have a tangent at $(x_0, f(x_0))$.*

• Question: which is the *equation* of that straight line?

• Answer: $y = ax + b$, where a and b can easily be determined based on two facts:

- As the slope of the straight line is given by a, it must be $a = f'(x_0)$

- Given that $y = f'(x_0)x + b$ crosses point $(x_0, f(x_0))$, it must be $f(x_0) = x_0 f'(x_0) + b$, i.e., $b = f(x_0) - x_0 f'(x_0)$

• Thus, *the straight line tangent to the graph of x_0 at $(x_0, f(x_0))$* is:

$$y = f'(x_0)x + f(x_0) - x_0 f'(x_0) = f(x_0) + (x - x_0)f'(x_0)$$

• $f(x_0) + (x - x_0)f'(x_0)$ is called the **linear or first order approximation of $f(x)$ at x_0** because for values of x close to x_0 we have:

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0)$$

• The approximation is made at x_0 and, hence, the function and the derivative are evaluated at x_0 in the approximation $\longrightarrow$ we look for a straight line that goes through $(x_0, f(x_0))$ and that has the same slope as the function f at that point.

4.6 Second derivative and second order approximation of a function

Second derivative

• When a function $y = f(x)$ is derivable at any point, the function derivative $f'(x)$ itself may be derivable. In this case we denote:

$$\frac{d(f'(x))}{dx} \text{ by } \frac{d^2y}{dx^2} \text{ or by } f''(x)$$

and that function is *called* **the second derivative** of $f(x)$.

• Interpretation: The *second derivative* at a point $f''(x)$ informs about the *impact on the first derivative* $f'(x)$ *of changes in* x. If $f''(x) > 0$ a slight *increase* of x entails a slight *increase* of $f'(x)$, and a slight decrease of x entails a slight decrease of $f'(x)$. If $f''(x) < 0$ a slight *increase* of x entails a slight *decrease* of $f'(x)$, and a slight decrease of x entails a slight increase of $f'(x)$.

Remark 1: The *sign* of $f''(x)$ can be interpreted in graphical terms and related to the **convexity** or **concavity** of $f(x)$'s graph. It may be shown that f concave (strictly concave) in a convex set D if and only if $f''(x) \leq 0$ ($f''(x) < 0$) for all $x \in D$ and that f convex (strictly convex) in a convex set D if and only if $f''(x) \geq 0$ ($f''(x) > 0$) for all $x \in D$.

Second order approximation of a function

• When $f(x)$ has second derivative at x_0 a better approximation of the function than the linear one (for values of x close to x_0) is possible using a quadratic function of the form $y = ax^2 + bx + c$, where a, b and c can be easily determined by three conditions:

$$\begin{gathered} ax_0^2 + bx_0 + c = f(x_0) \\ 2ax_0 + b = f'(x_0) \\ 2a = f''(x_0) \end{gathered}$$

Hence, it is obtained that:

$$\begin{gathered} a = \tfrac{f''(x_0)}{2} \\ b = f'(x_0) - x_0 f''(x_0) \\ c = f(x_0) - x_0 f'(x_0) + x_0^2 \tfrac{f''(x_0)}{2} \end{gathered}$$

and $y = ax^2 + bx + c = f(x_0) + (x - x_0)f'(x_0) + (x - x_0)^2 \frac{f''(x_0)}{2}$.

• The expression $f(x_0) + (x - x_0)f'(x_0) + (x - x_0)^2 \frac{f''(x_0)}{2}$ is called the **second order approximation of** $f(x)$ **at** x_0 because for values of x close to x_0 we have:

$$f(x) \approx f(x_0) + (x - x_0)f'(x_0) + (x - x_0)^2 \frac{f''(x_0)}{2}$$

4.7 Higher order derivatives and higher order approximations of a function

Higher order derivatives

• The n-order derivative of $f(x)$ is denoted by $f^n(x)$ and we have: $f^n(x) = \frac{d(f^{n-1}(x))}{dx}$.

Higher order approximation of a function:

- **Taylor's formula**

If f is $n + 1$ times differentiable (derivable: it has derivatives until the order $n + 1$) in an interval including x_0 and x then:

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + ... + \frac{f^n(x_0)}{n!}(x - x_0)^n + \frac{f^{n+1}(c)}{(n+1)!}(x - x_0)^{n+1}$$

where c is some number between x_0 and x.

Taylor's formula is used to approximate $f(x)$ at $x = x_0$ using the following polynomial of degree n:

$$f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + ... + \frac{f^n(x_0)}{n!}(x - x_0)^n$$

The order of the approximation is n. The term $\frac{f^{n+1}(c)}{(n+1)!}(x - x_0)^{n+1}$ is called the remainder or error if we use an approximation of order n for $f(x)$ at $x = x_0$. That error term may be used to obtain a bound of the error of the approximation (the highest value that the remainder may attain).

- **McLaurin's formula**

For $x_0 = 0$ the McLaurin's formula is obtained from the Taylor's formula:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + ... + \frac{f^n(0)}{n!}x^n + \frac{f^{n+1}(c)}{(n+1)!}x^{n+1}$$

where c is a number between 0 and x.

4.8 Global maximum and global minimum. Weierstrass' theorem.

Definition: If $y = f(x)$, where $x \in D \subset R$ (i.e., given $f : D \longrightarrow R$):

• $f(x)$ has a **global minimum on** D at $x_0 \in D$ if the value of $f(x)$ at x_0, i.e. $f(x_0)$, is smaller or equal than at *any other point of* D, and we write:

$$f(x_0) = \min_{x \in D} f(x) \text{ (minimum value of } f(x) \text{ on } D)$$

• $f(x)$ has a **global maximum on** D at $x_0 \in D$ if the value of $f(x)$ at x_0, i.e. $f(x_0)$, is greater or equal than at *any other point of* D, and we write:

$$f(x_0) = \max_{x \in D} f(x) \text{ (maximum value of } f(x) \text{ on } D)$$

• In both cases we say that $f(x)$ has a **global extreme** on D at x_0.

• **Remark 2**: Not always a function $y = f(x)$ has a global maximum or a global minimum on a set D. For instance, $1/x$ has a global minimum, but not a global maximum on $(0, 1]$.

• The following important theorem establishes sufficient conditions for $f(x)$ and D that ensure the existence of both global maximum and global minimum:

Weierstrass' Theorem: *If $y = f(x)$ is **continuous on a closed and bounded interval** $[a, b]$ (i.e., $f : [a, b] \longrightarrow R$ is continuous), then $f(x)$ has at least a global minimum and at least a global a maximum on $[a, b]$.*

4.9 Local maximum and local minimum. Conditions for a local extreme.

Definition: If $y = f(x)$, where $x \in D \subset R$ (i.e., given $f : D \longrightarrow R$):

• *$f(x)$ has a* ***local minimum on*** *D at $x_0 \in D$ if there is a neighbourhood B of x_0 such that the value of $f(x)$ at x_0 , i.e. $f(x_0)$, is smaller or equal than at any point $x \in B$ such that $x \neq x_0$*

• *$f(x)$ has a* ***local maximum*** *on D at $x_0 \in D$ if there is a neighbourhood B of x_0 such that the value of $f(x)$ at x_0 , i.e. $f(x_0)$, is greater or equal than at any point $x \in B$ such that $x \neq x_0$.*

• In both cases we say that $f(x)$ has a **local extreme** on D at x_0.

First-order condition for a local extreme:

• *An interior point of a set D is a point in that set for which there is a neighbourhood contained in set D* (examples: i) all points in $(1, 3)$ are interior points of $[1, 3]$, but 1 and 3 and are *not* interior points of that set and ii) the set $\left\{\frac{1}{n} : n \in N\right\} \subset R$ has *no* interior points).

Theorem: *If $y = f(x)$ is* ***derivable at an interior point*** *x_0 of a set D, and has a* ***local*** *extreme at x_0, then necessarily $f'(x_0) = 0$.*

• Equation $f'(x_0) = 0$ is referred to as "*first-order condition*" or "*necessary condition*" for an interior point to be a local extreme.

Second-order conditions for a local extreme:

• The first-order condition is in fact a *necessary* condition for an interior point to be a local extreme of a derivable function, but it is not sufficient, i.e., it may well be the case that the first-order condition holds but there is *no* local extreme.

• The following result establishes *sufficient* conditions for a local extreme for a function with first and second derivatives:

Theorem: *If* $y = f(x)$ *is* ***twice*** *derivable on* D *and if at an interior point* x_0 *of* D *it is*:

$$f'(x_0) = 0 \text{ and } f''(x_0) > 0 \text{ (resp. } f''(x_0) < 0) \tag{1}$$

then f(x) has a local minimum (resp. maximum) at x_0.

• Equations (1) are referred to as "*second-order conditions*" or "*sufficient conditions*" for an interior point to be a local extreme.

5 Integral calculus with one variable

5.1 Indefinite integrals

$\int f(x)dx = F(x) + c$ for any real number c, where $F'(x) = f(x)$.

Properties:

i) $\int (f(x) + g(x))dx = \int f(x)dx + \int g(x)dx$.
ii) $\int af(x)dx = a \int f(x)dx$.

Some important integrals:

i) $\int f(x)^a f'(x)dx = \frac{(f(x))^{a+1}}{a+1} + c$ (if $a \neq -1$) ($\Rightarrow \int x^a dx = \frac{x^{a+1}}{a+1} + c$).

ii) $\int \frac{f'(x)}{f(x)}dx = \ln(f(x)) + c$ ($\Rightarrow \int \frac{1}{x}dx = \ln(x) + c$).

iii) $\int e^{f(x)} f'(x)dx = e^{f(x)} + c$ ($\Rightarrow \int e^{ax}dx = \frac{e^{ax}}{a} + c$).

iv) $\int a^{f(x)} f'(x)dx = \frac{a^{f(x)}}{\ln(a)} + c$ ($\Rightarrow \int a^x dx = \frac{a^x}{\ln(a)} + c$).

5.2 Integration by parts

Since $\frac{d(f(x)g(x))}{dx} = (f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$ it follows that:

$$\int f(x)g'(x)dx = \int (f(x)g(x))'dx - \int f'(x)g(x)dx$$

$$= f(x)g(x) - \int f'(x)g(x)dx.$$

Example: $\int x \ln(x)dx$

Considering that $f(x) = \ln(x)$ and $g'(x) = x$ it follows that $f'(x) = \frac{1}{x}$ and $g(x) = \frac{x^2}{2}$. Hence:

$$\int x \ln(x) dx = \ln(x)\frac{x^2}{2} - \int \frac{x^2}{2}\frac{1}{x} dx = \ln(x)\frac{x^2}{2} - \frac{x^2}{4} + c.$$

5.3 Integration by substitution or change of variable

$\int f(g(x))g'(x)dx = \int f(u)du$ if $u = g(x)$ as then $du = g'(x)dx$.

Example: $\int \frac{1+\ln(x)}{x} dx$
$u = 1 + \ln(x) \Rightarrow du = \frac{1}{x}dx$
$\int \frac{1+\ln(x)}{x} dx = \int u du = \frac{u^2}{2} + c = \frac{1}{2}(1 + \ln(x))^2 + c.$

5.4 Definite integrals

Consider that f is bounded in $[a, b]$ (*a real valued function* f *of one variable is bounded in* $[a, b]$ *if there exists* $k \in R$ *such that* $-k \leq f(x) \leq k$ *for all* $x \in [a, b]$)

Proposition 1: If f is continuous in $[a, b]$ or if it is discontinuous only in a finite number of points in $[a, b]$ then f is (Riemann) integrable in $[a, b]$.

Remark 1: remember from Section 4.3 that f can be continuous and not derivable.

Barrow's rule: If f is continuous then:

$$\int_a^b f(x)dx = F(b) - F(a), \text{where } F(x) + c = \int f(x)dx$$

Example: $\int_1^3 \frac{3x}{10}dx$ (solution: $\frac{6}{5}$).

If $f \geq 0$ then $\int_a^b f(x)dx$ is equal to the area between the curve that represents f and the horizontal axis.

If $f \leq 0$ then $\int_a^b f(x)dx$ is a negative number.

Properties of definite integrals:

i) $\int_a^b f(x)dx = -\int_b^a f(x)dx$.

ii) $\int_a^a f(x)dx = 0$.

iii) $\int_a^b \alpha f(x)dx = \alpha \int_a^b f(x)dx$.

iv) $\int_a^b (f(x) + g(x))dx = \int_a^b f(x)dx + \int_a^b g(x)dx$.

v) $\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$ for any c such that $a < c < b$.

Theorem of the integral mean value: If f is continuous in $[a, b]$ and $a < b$ there exists θ such that $a < \theta < b$ and:

$$f(\theta) = \frac{1}{b-a} \int_a^b f(x)dx$$

Proposition 2: If f is continuous (Riemann integrable) in $[a, b]$ then $\varphi(x) = \int_a^x f(x)dx$ and $\phi(x) = \int_x^b f(x)dx$ are derivable (continuous) for all $x \in (a, b)$. Moreover, $\varphi'(x) = \phi'(x) = f(x)$ for all $x \in (a, b)$.

Remark 2: Note that a stronger result is obtained (φ and ϕ derivable, rather than only continuous) when the assumption on f is stronger (f continuous, rather than only Riemann integrable).

Integration by parts in the definite integral:

$$\int_a^b f(x)g'(x)dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)dx$$

Derivative with respect to another variable in the definite integral:

$$\frac{d}{dt}(\int_{a(t)}^{b(t)} f(x)dx) = f(b(t))b'(t) - f(a(t))a'(t)$$

$$\rightarrow example: \frac{d}{dt}(\int_{t}^{3} e^{-x^2}dx) = -e^{-t^2}$$

and:

$$\frac{d}{dt}(\int_{a(t)}^{b(t)} f(x,t)dx) = f(b(t),t)b'(t) - f(a(t),t)a'(t) + \int_{a(t)}^{b(t)} \frac{\partial f(x,t)}{\partial t}dx.$$

5.5 Introduction to differential equations

$x = x(t)$ with t: time ($t \geq 0$)

Examples:

i) $\frac{dx}{dt} = a$ where a constant
$\Rightarrow dx = adt \Rightarrow \int dx = \int adt \Rightarrow x(t) = at + c$

ii) $\frac{dx}{dt} = ax$ where a constant
$\Rightarrow \frac{dx}{x} = adt \Rightarrow \int \frac{dx}{x} = \int adt \Rightarrow \ln(x) = at + \ln(c) = \ln(ce^{at}) \Rightarrow x(t) = ce^{at}$, with $c > 0$ as the logarithm is not defined for negative numbers
$\rightarrow$ if $x(0) = x_0$ then $x(0) = ce^{a(0)} = c = x_0$ and $x(t) = x_0e^{at}$

iii) $\frac{dx}{dt} = a(k-x)$ where a and k are constants
$\Rightarrow -(\frac{-dx}{k-x}) = adt \Rightarrow -\int \frac{-dx}{k-x} = \int adt \Rightarrow -\ln(k-x) = at + \ln c \Rightarrow \ln(\frac{1}{k-x}) = \ln(ce^{at}) \Rightarrow \frac{1}{k-x} = ce^{at} \Rightarrow x(t) = k - \frac{1}{ce^{at}}$, with $k > x$ and $c > 0$ as the logarithm is not defined for negative numbers.

6 Real-valued functions of several variables

6.1 Introduction

Real-valued function of n variables: $f(\mathbf{x}) = f(x_1, x_2, x_3, ..., x_n)$, $f : D \subset R^n \to R$.

Level set of value k for function f: $S_k = \{\mathbf{x} \in D : f(\mathbf{x}) = k\}$.

6.2 Continuity

Continuity of $f(\mathbf{x})$ at point $\mathbf{x}_0 = (x_{10}, x_{20}, ..., x_{n0}) \in \mathbf{R}^n$: analogous definition to the case of one variable (properties of continuity also analogous) $\to$ but it is necessary to consider a neighborhood of $f(\mathbf{x}_0)$ and an open ball around $\mathbf{x}_0$ $\longrightarrow$ an open ball (a neighbourhood) around $\mathbf{x}_0$.

Limit of $f(\mathbf{x})$ at point $\mathbf{x}_0 \in \mathbf{R}^n$: existence requires to consider all trajectories that approach $\mathbf{x}_0$ $\to$ otherwise the relationship between limit and continuity as in the case of one variable. Properties of limits analogous to the case of one variable.

6.3 Partial derivatives. Gradient. Hessian matrix.

Definition: **Partial derivative** of $f(\mathbf{x})$ at $\mathbf{x}_0$ with respect to x_i:

$$\frac{\partial f}{\partial x_i}(\mathbf{x}_0) = \lim_{h \to 0} \frac{f(x_{10}, x_{20}, ..., x_{i0} + h, ..., x_{n0}) - f(x_{10}, x_{20}, ..., x_{i0}, ..., x_{n0})}{h}$$

Example:

$f(x_1, x_2) = x_1 x_2$ and $\mathbf{x}_0 = (x_{10}, x_{20}) = (1, 3)$

$\frac{\partial f(x_1,x_2)}{\partial x_1} = x_2$ and $\frac{\partial f(x_1,x_2)}{\partial x_2} = x_1$
at $\mathbf{x}_0$: $\frac{\partial f}{\partial x_1} = 3$ and $\frac{\partial f}{\partial x_2} = 1$

Definition: Gradient of $f(\mathbf{x})$ at $\mathbf{x}_0$ is: $\nabla f := (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}) \in \mathbf{R}^n$, with all partial derivatives evaluated at $\mathbf{x}_0$.

Example: In what points of function $f(x_1, x_2) = 4x_1^2 + x_2^2 - 4x_1x_2$ is its gradient 0, that is, $\nabla f(x_1, x_2) = (0, 0)$? $\longrightarrow$ Both partial derivatives become 0 when $x_2 = 2x_1$.

Second order partial derivatives: $\frac{\partial}{\partial x_j}(\frac{\partial f}{\partial x_i}) = \frac{\partial^2 f}{\partial x_i \partial x_j}$

Definition: **Hessian matrix**: second order derivatives matrix:

$$H(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Example: The Hessian matrix for function $f(x_1, x_2) = 4x_1^2 + x_2^2 - 4x_1x_2$ is:

$$H(f) = \begin{pmatrix} 8 & -4 \\ -4 & 2 \end{pmatrix}$$

Remark 1: The Hessian matrix in this example does not depend on x_1 and x_2. However, in general, the second order partial derivatives and, hence, the Hessian matrix may depend on the levels of x_i.

Proposition: If the second derivatives are continuous functions then $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ (the Hessian matrix is symmetric).

6.4 Differentiability. Total differential.

Intuition for differentiability: f differentiable at $\mathbf{x}_0 \in R^n$ if it is "smooth" at an open ball centered at $\mathbf{x}_0$ for all trajectories on f through that point.

Propositions:

i) f differentiable at $\mathbf{x}_0 \Rightarrow f$ continuous at $\mathbf{x}_0$.

ii) f differentiable at $\mathbf{x}_0 \Rightarrow$ the partial derivatives of f with respect to each component of $\mathbf{x}$ exist at $\mathbf{x} = \mathbf{x}_0$ ($\frac{\partial f}{\partial x_i}(\mathbf{x}_0)$ exists for all $i = 1, 2, ..., n$). The contrary is not true.

iii) $\frac{\partial f}{\partial x_i}(\mathbf{x}_0)$ exists and is continuous in an open ball centered at $\mathbf{x}_0$ for all $i = 1, 2, ..., n \Rightarrow f$ differentiable at $\mathbf{x}_0$.

Definition: The **total differential** of $f(\mathbf{x})$ at $\mathbf{x}_0$ is: $df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i$, with all partial derivatives evaluated at $\mathbf{x}_0$. The total differential captures the impact on f of infinitesimal simultaneous variations in some or all components of $\mathbf{x}$.

6.5 The chain rule

Case i) If $\mathbf{z} = f(x, y)$ with $x = h(t)$ and $y = g(t)$ then:

$$\frac{dz}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

Example:
$z = xe^{2y}$ with $x = \sqrt{t}$ and $y = \ln(t)$ (solution: $\frac{dz}{dt} = \frac{t^2}{2\sqrt{t}} + 2t\sqrt{t}$)

Case ii) If $z = f(x, y)$ with $x = h(t, s)$ and $y = g(t, s)$ then:

$$\frac{\partial z}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$$
$$\frac{\partial z}{\partial s} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial s}$$

Example:
$z = x^2 + 2y^2$ with $x = t - s^2$ and $y = ts$ (solution: $\frac{\partial z}{\partial t} = 2t - 2s^2 + 4ts^2$ and $\frac{\partial z}{\partial s} = 4t^2 s - 4ts + 4s^3 \rightarrow$ the derivatives for particular values of t and s may be obtained).

6.6 Concavity and convexity of real-valued functions of several variables

Definition: *Let f be a real-valued function defined on a convex set X in R^n. The function f is called a:*

i) ***concave function*** *if, for all $\mathbf{x}, \mathbf{y} \in X$ and $0 \leq \theta \leq 1$, $f[\theta\mathbf{x} + (1-\theta)\mathbf{y}] \geq \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y})$*

ii) ***strictly concave function*** *if, for all $\mathbf{x}, \mathbf{y} \in X$, with $\mathbf{x} \neq \mathbf{y}$, and $0 < \theta < 1$,$f[\theta\mathbf{x} + (1-\theta)\mathbf{y}] > \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y})$*

iii) ***convex function*** *if, for all $\mathbf{x}, \mathbf{y} \in X$ and $0 \leq \theta \leq 1$, $f[\theta\mathbf{x} + (1-\theta)\mathbf{y}] \leq \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y})$*

iv) ***strictly convex function*** *if, for all $\mathbf{x}, \mathbf{y} \in X$, with $\mathbf{x} \neq \mathbf{y}$, and $0 < \theta < 1$,*

$$f[\theta\mathbf{x} + (1-\theta)\mathbf{y}] < \theta f(\mathbf{x}) + (1-\theta)f(\mathbf{y})$$

Theorem: *Let $f(\mathbf{x})$ be twice continuously differentiable real-valued function on an open convex set X in R^n. Then:*

i) The function f is convex on X if and only if for all $\mathbf{x} \in X$ all leading principal minors of the Hessian matrix are greater or equal than 0,

ii) The function f is strictly convex on X if and only if for all $\mathbf{x} \in X$ all leading principal minors of the Hessian matrix are positive,

iii) The function f is concave on X if and only if for all $\mathbf{x} \in X$ leading principal minors of the Hessian matrix alternate in sign in the following way: leading principal minor of order 1 is ≤ 0, leading principal minor of order 2 is ≥ 0, leading principal minor of order 3 is ≤ 0,...., and

iv) The function f is strictly concave on X if and only if for all $\mathbf{x} \in X$ leading principal minors of the Hessian matrix alternate in sign starting with the negative sign (leading principal minor of order 1 is negative, leading principal minor of order 2 is positive, leading principal minor of order 3 is negative,...).

Example: Check the convexity/concavity of the function $f(x_1, x_2) = (x_1 - 2)^3 + 2x_1x_2 + x_2^2$ on the convex set $D = \{(x_1, x_2) : x_1 > 3\} \longrightarrow$ Solution:

Since $\frac{\partial f}{\partial x_1} = 3(x_1-2)^2 + 2x_2$ and $\frac{\partial f}{\partial x_2} = 2x_1 + 2x_2$ it follows that:

$$H(f) = \begin{pmatrix} 6(x_1-2) & 2 \\ 2 & 2 \end{pmatrix}.$$

Hence, leading principal minor (l.p.m.) of order 1 is $6(x_1-2)$ and l.p.m. of order 2 is $12(x_1-2)-4 = 12x_1 - 28 = 4(3x_1-7)$. When $x_1 > 3$ both l.p.m. are positive. Therefore, f is strictly convex.

Remark 2: Log concavity of f ($\log(f)$ is concave) is a weaker requirement than concavity of f.

6.7 Approximations of a function of several variables

The n-order approximation of a real valued function of several variables at a point may be obtained as in Sections 4.5, 4.6 and 4.7. For instance the second-order approximation of a real valued function of several variables at $\mathbf{x}_0$ is:

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \left.\frac{\partial f(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\mathbf{x}_0} (x_i - x_{i0}) + \frac{1}{2!}\sum_{i=1}^{n}\left[\sum_{j=1}^{n} \left.\frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0} (x_j - x_{j0})\right](x_i - x_{i0})$$

6.8 Homogeneous functions. Euler's theorem

Definition: f is **homogeneous of degree** k if $f(ax_1, ax_2, ..., ax_n) = a^k f(x_1, x_2, ..., x_n)$ for all $a \in R$ and $a > 0$.

Examples:
i) $f(x_1, x_2) = 3x_1^2 + 2x_1x_2 + x_2^2$ is homogeneous of degree 2.
ii) $f(x_1, x_2) = \frac{x_1+x_2}{3x_1-x_2}$ is homogeneous of degree 0.

iii) $f(K,L) = AK^{\alpha}L^{1-\alpha}$ is homogeneous of degree 1.

Proposition: If f is homogeneous of degree k then $\frac{\partial f}{\partial x_i}$ is homogeneous of degree $k-1$ for all $i = 1, 2..., n$.

Proof:

$$f(ax_1, ax_2, ..., ax_i, ..., ax_n) = a^k f(x_1, x_2, ..., x_i, ..., x_n)$$

$$\Rightarrow a\frac{\partial f(ax_1,ax_2,...,ax_i,...,ax_n)}{\partial x_i} = a^k \frac{\partial f(x_1,x_2,...,x_i,...,x_n)}{\partial x_i}$$

$$\Rightarrow \frac{\partial f(ax_1,ax_2,...,ax_i,...,ax_n)}{\partial x_i} = a^{k-1} \frac{\partial f(x_1,x_2,...,x_i,...,x_n)}{\partial x_i}$$

Euler's theorem: If f differentiable and $\mathbf{x}, \mathbf{ax} \in D$ then:

$$f \text{ is homogeneous of degree } k \text{ in } D$$
$$\Leftrightarrow$$
$$x_1\frac{\partial f}{\partial x_1} + x_2\frac{\partial f}{\partial x_2} + ... + x_n\frac{\partial f}{\partial x_n} = kf(x_1, x_2, ..., x_n)$$

6.9 Multiple integrals

Double integrals over domains in R^2:$\int_a^b \left[\int_c^d f(x_1, x_2)dx_2\right] dx_1$ (x_1 takes values between a and b and x_2 takes values between c and d, where $a, b, c, d \in R$).

If $f(x_1, x_2) \geq 0$ for all x_1 and x_2 such that $a \leq x_1 \leq b$ and $c \leq x_2 \leq d$ then $\int_a^b \left[\int_c^d f(x_1, x_2)dx_2\right] dx_1$ is equal to the volume below the surface that represents f and above the square in R^2 of points (x_1, x_2) such that $a \leq x_1 \leq b$ and $c \leq x_2 \leq d$.

If x_1 takes values between a and b and x_2 takes values between c and d, where $a, b, c, d \in R$ then:

$$\int_a^b \left[\int_c^d f(x_1, x_2)dx_2\right] dx_1 = \int_c^d \left[\int_a^b f(x_1, x_2)dx_1\right] dx_2$$

The result obtained when the integral on x_2 is considered first is the same as the result obtained when the integral on x_1 is considered first.

However, the integral on x_2 has to be considered first in $\int_a^b \left[\int_{h(x_1)}^{g(x_1)} f(x_1, x_2)dx_2\right] dx_1$, where $h(x_1) < g(x_1)$ for $a \leq x_1 \leq b$. That integral is the volume in $\mathbf{R}^3$ under the function $f(x_1, x_2)$ and above the area delimited by the functions $x_2 = h(x_1)$ and $x_2 = g(x_1)$ for $a \leq x_1 \leq b$ in the plane defined by the x_1-axis and the x_2-axis.

6.10 Appendix: Theorem of the implicit function and implicit differentiation

Theorem of the Implicit Function: *Let f be a continuous real valued function defined on $D \subset R^n$ such that f has continuous partial derivatives in an open ball of centre $\mathbf{x}_0 \in D$. If $f(\mathbf{x}_0) = c$ and $\frac{\partial f}{\partial x_i}(\mathbf{x}_0) \neq 0$, for $1 \leq i \leq n$, then there exists $h(x_1, x_2, ..., , x_{i-1}, x_{i+1,...}x_n)$ and an open ball of x_0 such that $h(x_{10}, x_{20}, ..., , x_{(i-1)0}, x_{(i+1)0}, ..., x_{n0}) = x_{i0}$, $f(x_1, x_2, ..., x_{i-1}, h(x_1, x_2, ..., , x_{i-1}, x_{i+1,...}x_n), x_{i+1}, ..., x_n) = c$ for any $\mathbf{x}$ in that open ball and h is derivable with continuous derivatives in that ball.*

Implicit function´s rule: Consider that $f(x_1, x_2, ..., , x_i, ..., x_n) = c$ and there exists a function $x_i = h(x_1, x_2, ..., , x_{i-1}, x_{i+1,...}x_n)$ defined implicitly. If we derive with respect to x_j both sides of $f(x_1, x_2, ..., , x_i, ..., x_n) = c$ it follows that:

$$\frac{df}{dx_j} = \frac{\partial f}{\partial x_j} + \frac{\partial f}{\partial x_i}\frac{\partial h}{\partial x_j} = 0 \Rightarrow \frac{\partial h}{\partial x_j} = -\frac{\frac{\partial f}{\partial x_j}}{\frac{\partial f}{\partial x_i}} \qquad ((1))$$

(the partial derivatives of h with respect to each variable can be obtained without knowing the function h!; it is enough to know that h exists)

Example:

Consider that $x_1 - 2x_2 - 3x_3 + (x_3)^2 = -2$. The function $f(x_1, x_2, x_3) = x_1 - 2x_2 - 3x_3 + (x_3)^2$ is continuous and its partial derivatives ($\frac{\partial f}{\partial x_1} = 1$, $\frac{\partial f}{\partial x_2} = -2$ and $\frac{\partial f}{\partial x_3} = -3 + 2x_3$) are also continuous. Then there exists x_3 as an implicit function of x_1 and x_2, that function is derivable with continuous derivatives in a ball of any point (x_1, x_2, x_3) in which $x_1 - 2x_2 - 3x_3 + (x_3)^2 = -2$ and the partial derivatives of x_3 with respect to x_1 and with respect to x_2 for the points in that ball may be obtained from (1): $\frac{\partial x_3}{\partial x_1} = -\frac{1}{-3+2x_3} = \frac{1}{3-2x_3}$ and $\frac{\partial x_3}{\partial x_2} = -\frac{-2}{-3+2x_3} = \frac{-2}{3-2x_3}$.

$\frac{\partial x_3}{\partial x_1}$ and $\frac{\partial x_3}{\partial x_2}$ may also be obtained by writing $x_3 = h(x_1, x_2)$ and by differentiating directly the function f with respect to x_1 and to x_2, respectively:

$$1 - 3\frac{\partial x_3}{\partial x_1} + 2x_3\frac{\partial x_3}{\partial x_1} = 0 \Rightarrow \frac{\partial x_3}{\partial x_1} = \frac{1}{3-2x_3} \qquad (A)$$

$$-2 - 3\frac{\partial x_3}{\partial x_2} + 2x_3\frac{\partial x_3}{\partial x_2} = 0 \Rightarrow \frac{\partial x_3}{\partial x_2} = \frac{-2}{3-2x_3} \qquad (B)$$

To obtain $\frac{\partial^2 x_3}{\partial x_1^2}$ we differentiate equation (A) with respect to x_1: $-3\frac{\partial^2 x_3}{\partial x_1^2} + 2(\frac{\partial x_3}{\partial x_1})^2 + 2z\frac{\partial^2 x_3}{\partial x_1^2} = 0 \Rightarrow \frac{\partial^2 x_3}{\partial x_1^2} = \frac{2(\frac{\partial x_3}{\partial x_1})^2}{3-2x_3} = \frac{2(\frac{1}{3-2x_3})^2}{3-2x_3} = \frac{2}{(3-2x_3)^3}$.

To obtain $\frac{\partial^2 x_3}{\partial x_1 \partial x_2}$ we differentiate equation (A) with respect to x_2 (or equation (B) with respect to x_1) $-3\frac{\partial^2 x_3}{\partial x_1 \partial x_2} + 2\frac{\partial x_3}{\partial x_1}\frac{\partial x_3}{\partial x_2} + 2z\frac{\partial^2 x_3}{\partial x_1 \partial x_2} = 0 \Rightarrow \frac{\partial^2 x_3}{\partial x_1 \partial x_2} = \frac{2\frac{\partial x_3}{\partial x_1}\frac{\partial x_3}{\partial x_2}}{3-2x_3} = \frac{2\frac{1}{3-2x_3}\frac{-2}{3-2x_3}}{3-2x_3} = \frac{-4}{(3-2x_3)^3}$.

To obtain $\frac{\partial^2 x_3}{\partial x_2^2}$ we differentiate equation (B) with respect to x_2: $-3\frac{\partial^2 x_3}{\partial x_2^2} + 2(\frac{\partial x_3}{\partial x_2})^2 + 2z\frac{\partial^2 x_3}{\partial x_2^2} = 0 \Rightarrow \frac{\partial^2 x_3}{\partial x_2^2} = \frac{2(\frac{\partial x_3}{\partial x_2})^2}{3-2x_3} = \frac{2(\frac{-2}{3-2x_3})^2}{3-2x_3} = \frac{8}{(3-2x_3)^3}$.

7 Functions of R^n in R^m

Consider a function f of R^n into R^m: $f(x_1, x_2, ..., x_n) = (y_1, y_2, ..., y_m)$. Denote by f_i the function such that $y_i = f_i(x_1, x_2, ..., x_n)$ for $i = 1, 2, ..., m$.

7.1 The Jacobian matrix and the Jacobian determinant

The **Jacobian matrix** of f is:

$$J(f) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

When $m = n$ the **Jacobian determinant** of f is:

$$|J(f)| = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{vmatrix}$$

7.2 Composition of functions

Composition of functions:

Example: Let $f(x_1, x_2) = (x_1^2 + x_2^2, x_2, x_1)$ and $g(y_1, y_2, y_3) = (2y_1 + y_3, 4y_2 + y_3)$.

i) Obtain the composite map $g.f : R^2 \to R^2$

$(g.f)(x_1, x_2) = g(f(x_1, x_2)) = g(x_1^2 + x_2^2, x_2, x_1) = (2(x_1^2 + x_2^2) + x_1, 4x_2 + x_1)$

ii) Calculate the **Jacobian determinant** of $g.f$ at point $(1, 1)$

$$|J(g.f)| = \begin{vmatrix} 4x_1 + 1 & 4x_2 \\ 1 & 4 \end{vmatrix}$$

$$|J(g.f)|_{(1,1)} = \begin{vmatrix} 5 & 4 \\ 1 & 4 \end{vmatrix} = 16$$